How Emergent Self Organizing Maps can help counter domestic violence

Jonas Poelmans[1], Paul Elzinga[3], Stijn Viaene[1,2], Marc M. Van Hulle[5], Guido Dedene[1,4]

[1]K.U.Leuven, Faculty of Business and Economics, Naamsestraat 69,
3000 Leuven, Belgium
[2]Vlerick Leuven Gent Management School, Vlamingenstraat 83,
3000 Leuven, Belgium
[3]Police Organisation Amsterdam-Amstelland, James Wattstraat 84,
1000 CG Amsterdam, The Netherlands
[4]Universiteit van Amsterdam Business School, Roetersstraat 11
1018 WB Amsterdam, The Netherlands
[5]K.U.Leuven, Laboratorium voor Neuro- en Psychofysiologie
Campus Gasthuisberg O&N2, Bus 1021, Herestraat 49
3000 Leuven, Belgium
{Jonas.Poelmans, Stijn.Viaene, Guido.Dedene}@econ.kuleuven.be
Paul.Elzinga@amsterdam.politie.nl
marc@neuro.kuleuven.be

Abstract

Topographic maps are an appealing exploratory instrument for discovering new knowledge from databases. During the past years, new types of Self Organizing Maps (SOM) were introduced in the literature, including the recent Emergent SOM. The ESOM is used to study a large set of police reports describing a whole range of violent incidents that occurred during the year 2007 in the police region Amsterdam-Amstelland (the Netherlands). It is demonstrated that it provides an exploratory search instrument for examining unstructured text in police reports. First, it is shown how the ESOM was used to discover a whole range of new features that better distinguish domestic from non-domestic violence cases. Then, it is demonstrated how this resulted in a significant improvement in classification accuracy. Finally, the ESOM is showcased as a powerful instrument for the domain expert interested in an in-depth investigation of the nature and scope of domestic violence.

1. Introduction

In 1997, the ministry of Justice of the Netherlands made its first inquiry into the nature and scope of domestic violence [2]. It turned out, 45% of the population once fell victim to non-incidental domestic violence. For 27% of the population, the incidents even occurred on a weekly or daily basis. These gloomy statistics brought this topic to the center of the political agenda. By consequence, acting firmly against this phenomenon became one of the pivotal projects of the administration of Prime Minister Balkenende of the Netherlands. [1]

According to the department of Justice and the police organization of the Netherlands, domestic violence can be characterized as serious acts of violence committed by someone of the domestic sphere of the victim. Violence includes all forms of physical assault. The domestic sphere includes all partners, ex-partners, family members, relatives and family friends of the victim. Family friends are those persons who have a friendly relationship with the victim and who (regularly) meet the victim in his/her home [1].

Pursuing an effective policy against offenders is nowadays one of the top priorities of the police organization of the region Amsterdam-Amstelland in the Netherlands. Of course, in order to pursue an effective policy against offenders, being able to swiftly recognize cases of domestic violence and label reports accordingly, are of the utmost importance. Still this has proven to be problematic. In the past, intensive audits of the police databases related to filed reports have shown that many reports tended to be wrongly classified.

To cope with this problem, attempts have been made to develop a technique that automatically classifies cases as domestic or non-domestic violence. A multi-layer perceptron and an SVM were used, but

[1]http://www.regering.nl/Het_kabinet/Eerdere_kabinetten/Kabinet_Balkenende_II/Regeerakkoord#internelink4

unfortunately classification accuracy was around 80% only. Moreover, these techniques did not provide any insight in the performed classification, since they are black-boxes.

In the current paper, this problem will be tackled using a special class of topographic maps [4] called Emergent Self Organizing Maps (ESOM) [5], which are particularly suited for high-dimensional data visualization. It will be demonstrated that from the unstructured text in police reports, essential knowledge regarding domestic violence is obtained by using the ESOM. In addition, it will be shown that an efficient, comprehensible and highly accurate automated classification model can be constructed using an ESOM.

The remainder of this paper is as follows. In section 2, we shall cover the essentials of topographic map theory, and in particular the Emergent Self Organizing Maps. In section 3, the used dataset will be discussed; after which, in section 4, the ESOM application to the domestic violence problem is demonstrated. Finally, section 5 concludes the paper.

2. Topographic Map essentials

From a practitioner's point of view, topographic maps are a particularly appealing technique for knowledge discovery in databases [12]. It performs a non-linear mapping of the high-dimensional data space to a low-dimensional one, usually a two-dimensional one, which enables the visualization and exploration of the data [9]. It can be used to detect clusters and it maintains the neighborhood relationships that are present in the input space. It also provides the user with an idea of the complexity of the dataset, the distribution of the dataset (e.g. spherical) and the amount of overlap between the different classes. The lower-dimensional data representation is also an advantage when constructing classifiers. Finally, only a minimal amount of expert knowledge is required for the user to be able to use it effectively for exploratory data analysis.

2.1. Emergent SOM

An Emergent Self Organizing Map (ESOM) is a very recent type of topographic map [5]. It is argued to be especially useful for visualizing sparse, high-dimensional datasets, yielding an intuitive overview of its structure [7]. An Emergent SOM differs from a traditional SOM in that a very large number of neurons (at least a few thousands) are used [6]. It is said that the topology preservation of the traditional SOM projection is of little use when using small maps: the performance of a small SOM is argued to be almost identical to that of a *k*-means clustering, with *k* equal to the number of nodes in the map [5]. An additional advantage of an ESOM is that it can be trained directly on the available dataset without first having to perform a feature selection procedure [8]. ESOM maps can be created and used for data analysis by means of the publicly available *Databionics ESOM Tool* [16]. This tool allows the user to construct both flat and unbounded (i.e., toroidal) ESOM maps.

2.2. ESOM and domestic violence

The reason why ESOM was chosen is threefold. First, in the literature, the need for exploratory data analysis has often been described [14] The aims to achieve with exploratory search are: knowledge acquisition, discovering gaps in the existing knowledge, comprehension of concepts and the discovery of the boundaries of meaning for key concepts. Exploratory search requires strong human participation in a continuous and exploratory process. To effectively support the full range of search activities, humans should be brought more actively into the search process. This can only be achieved by tools that offer highly interactive user interfaces that continuously engage human control over the information seeking process [13].

In this paper, it is demonstrated that the ESOM is one of those rare tools that meet this key requirement. It is shown that it provides a highly interactive user interface that moves exploratory search process beyond predictable fact retrieval.

Second, in 2007, the database of the Amsterdam-Amstelland police contained more than 7000 cases that contained a statement made by the victim of a violent incident to the police. Because it is physically impossible for any individual to process this sheer amount of information, applying text mining technology seemed a natural approach. Text mining has been defined as "the discovery by computer of new, previously unknown, information by automatically extracting information from different written resources" [15]. A pivotal step in this process is the text analysis phase. This approach has been tried out in the past, but it turned out that the results were not convincing enough. This was for a large part due to the lack of a good thesaurus.

In this paper, it is demonstrated that the ESOM tool provides the ideal exploratory instrument for supporting the critical text analysis phase. Because of its highly interactive user interface, human participation and effective use of their expert prior knowledge in the search process is promoted.

Finally, when a victim of a violent incident makes a statement to the police, the police officer has to judge whether or not it is domestic violence. If he considers it to be domestic violence, he can assign the domestic violence label to the case. Because it is a very costly task to classify cases and to verify whether or not the performed classifications are correct, the introduction of an automated classifier would result in major savings. It is clear that, a high overall accuracy, a low false negative rate and a comprehensible classification are key requirements.

One of the most important advantages of a nearest neighbour classifier based on an Emergent Self Organizing Map is the comprehensibility of the performed classification. To answer the question why a police report was classified as domestic or as non-domestic violence, one simply needs to inspect the cases that belong to the best matched neuron(s) of the police report.

It shall be demonstrated that a nearest-neighbor classifier based on the ESOM meets all these requirements and outperforms other more complex classifiers such as the SVM, Naïve Bayes, multi-layer perceptron, etc.

3. Dataset

The database of the Amsterdam police organization contains all documents relating to criminal offences. Documents related to certain types of crimes receive corresponding labels. Immediately after the reporting of a crime, police officers are given the possibility to judge whether or not it is a domestic violence case. If they believe this is the case, they can assign the label domestic violence to the report. However, not all domestic violence cases are recognized as such by police officers and by consequence, many police reports are wrongly lacking a "domestic violence" label.

The dataset consists of a selection of 4814 police reports describing a whole range of violent incidents from the year 2007. All domestic violence cases from that period are a subset of this dataset. This selection came about amongst others by filtering out those police reports that did not contain the reporting of a crime by a victim, which is necessary for establishing domestic violence. This happens for example when a police officer was sent to an incident and later on wrote a report in which he/she mentioned his/her findings, while the victim did not make an official statement to the police. The follow-up reports referring to previous cases were also removed. From the 4814 police reports contained in the dataset, the person who reported the crime, the suspect, the persons involved in the crime, the witnesses, the project code and the statement made by the victim to the police were extracted. Of these 4814 reports, 1657 were cases of domestic violence; the others not. These data were used to generate the 4814 html-documents that were used during the research. An example of such a report is displayed in Fig. 1.

Title of incident	Violent incident xxx
Reporting date	31-03-2008
Project code	Domestic violence against ex-partner
Crime location	Amsterdam Wibautstraat yyy
Suspect (male) Suspect (18-45yr)	zzz
Address	Amsterdam Waterlooplein yyy
Involved (male) Involved (>45yr)	Neighbours
Address	Amsterdam Wibautstraat www
Victim (female) Victim (18-45yr)	uuu
Address	Amsterdam Waterlooplein vvv

Reporting of the crime

Yesterday morning I was taking a bath. Suddenly my daughter ran into the bathroom followed by her ex-boyfriend. She screamed for help. He had a gun in his hand and he was clearly under influence of beer or drugs. He yelled out that he couldn't live without her. He threatened to kill me and my daughter if she wouldn't come back to their house. The neighbours who were alarmed by all the noise came to give some help. Meanwhile another neighbour

phoned the police. I jumped out of my bath and tried to push him on the floor. During this fight I got some serious injuries on my back etc.

Fig. 1. Example police report.

The initial thesaurus – a collection of terms – was obtained by performing frequency analyses on these police reports. The relevant terms that occurred most often were retrieved and added to the initially empty thesaurus. This resulted in an initial set of 123 terms. In the dataset, for each police report, it is indicated which of these terms are present. An excerpt of this dataset is displayed in table 1.

Table 1. Excerpt of the categorical dataset used during the research.

	kicking	Dad hits me	Stabbing	cursing	scratching	maltreating
Report 1	X	X				X
Report 2			X	X	X	
Report 3	X	X	X	X	X	
Report 4						X
Report 5				X	X	

An initial analysis of the data revealed that the two cases (domestic and non-domestic violence) do not appear in separate clusters, which makes the use of a clustering followed by a labeling of the high density peaks not a viable approach (see Fig. 2 where the high density regions (darker pixels) do not correspond to separate labels). Topographic maps such as the ESOM may overcome this since they not use the labels but rather approximate the data manifold.

In a first step, an ESOM map with a toroidal topology of the neurons as well as a flat topology were trained using this dataset, in order to capture the distribution of the dataset. For the ESOM, the standard parameter settings of the Databionics software were used. We did not attempt to optimize them. A SOM with a lattice containing 50 rows and 82 columns of neurons was used (50x82=4100 neurons in total). The weights were initialized randomly by sampling a Gaussian with the same mean and standard deviation as the corresponding features. A Gaussian bell-shaped kernel with initial radius of 24 was used as a neighborhood function. Further, an initial learning rate of 0.5 and a linear cooling strategy for the learning rate were used. The number of training epochs was set to 20. In the map displayed in Fig. 2, the best matching (nearest-neighbor) nodes are labeled in the two classes for the given test data set (red for domestic violence, green for non-domestic violence). The red squares in all figures represent neurons that mainly contain domestic violence reports, whereas the green squares represent neurons that mainly contain non-domestic violence reports.

Analyzing the ESOM map, based on the categorical dataset, led us to conclude that it was spherically distributed. It can be seen that there is one large domestic violence cluster located at the center of the map, and a domestic violence cluster running upward and to the left of the map. The latter continues over the edge of the map (note that the map is actually toroidal) and has an outlier on the right of the map. Moreover, when the flat ESOM map was compared to the toroidal ESOM map, the toroidal map provided a much better visualization of the dataset. The border effect was clearly present in the flat map resulting in undesired distortions of the map. Most of the observed clusters were located at the border of the map, which made them smaller in area, and the large group of domestic violence cases was less clearly demarcated from the non-domestic violence cases. Therefore, it seemed more natural to use a toroidal ESOM for visualizing this dataset.

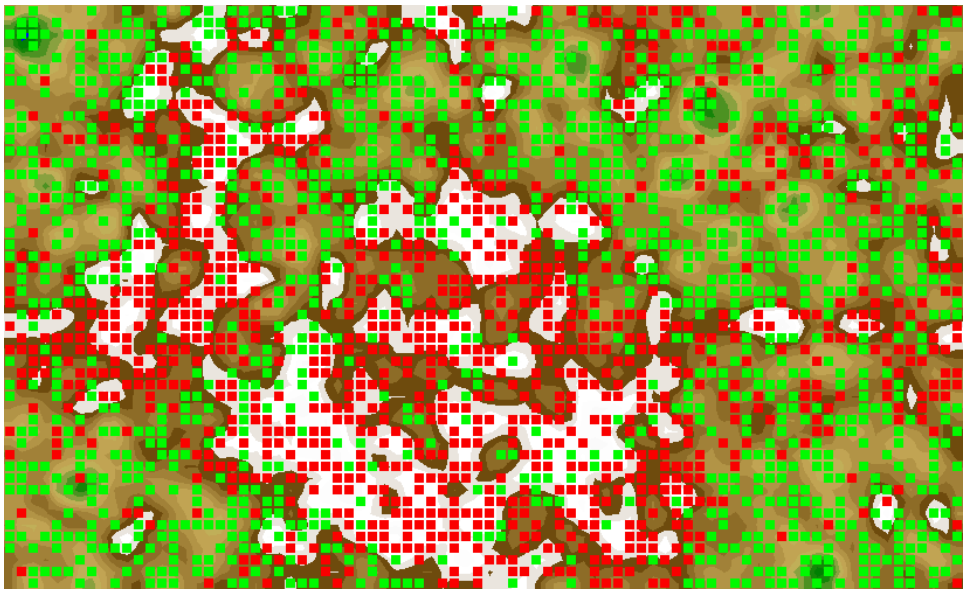

Fig. 2. Toroidal ESOM map trained on the categorical dataset with all features

The map displayed in Fig. 2 was trained directly on the entire dataset with 123 features. Fig. 2 clearly shows that the density profile of the ESOM map does not match the uniform distribution of the labeled data vectors. Moreover, there is no ridge in the map that separates the domestic- from the non-domestic violence cases. Therefore, the 'watershed' technique [6] will not lead to a correct identification of the classes. We have also explored that case where we chose to apply feature selection to reduce the input space dimensionality, prior applying ESOM. A heuristic feature selection procedure, known as minimal-redundancy-maximal-relevance (mRMR), as described in [11], was considered. The aim was to select the 50 most relevant features. To obtain the optimal feature set, an SVM, a Neural Network, a kNN (with k=3) and a Naïve Bayes classifier were used to measure the classification performance for an increasing number of features. The classification performance is plotted as a function of the number of features in Fig. 3.

It was chosen to retain the best 44 features. A toroidal ESOM map was trained on this dataset with a reduced number of features and was compared to that of Fig. 2. However, the density problem was not solved by lowering the number of features.

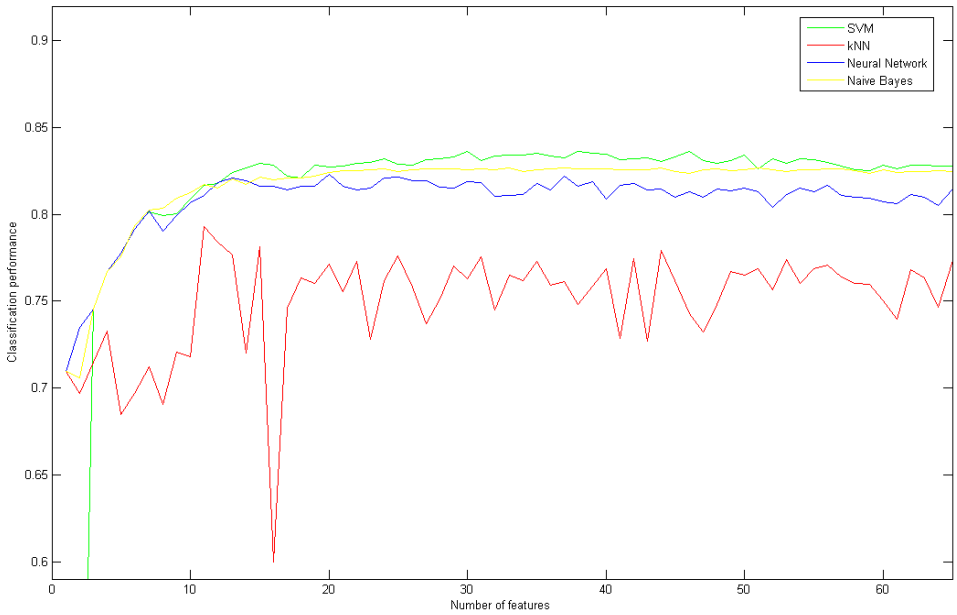

Fig. 3. Classification performance

Finally, a kNN classifier was built for the ESOM maps. For ESOM, k was set to 1 and 2 consecutively. In order to obtain the misclassification error of the ESOM map, the Euclidean distance of each input vector to each weight vector was measured. For each weight vector (corresponding to a node of the map) it was calculated how many of the domestic and non-domestic violence cases had this weight vector as a best match. If the node dominantly contained domestic violence cases, it was labeled as a domestic violence node and the non-domestic violence cases that best matched to this node were considered to be wrong classifications. A best-matched neuron is a neuron for which there exists at least one input vector for which the Euclidean distance to the weight vector of this node is minimal.

4. Results

An interesting result is that the map provides a relatively good portioning of domestic and non-domestic violence cases. Many of the best matched nodes dominantly contain either domestic or non-domestic violence cases. This indicates that there is only a limited amount of overlap between them. The observed overlap probably indicates that the feature set is not sufficiently refined to discriminate between the two classes. However, it should be noted that some cases might have been wrongly classified by the police officers.

It is interesting to observe that some red squares are located in the middle of a large group of green squares and vice versa. After an in-depth manual inspection of the police reports corresponding to these red outliers, some interesting discoveries were made. Surprisingly, only a small part of these police reports turned out to be incorrectly classified as domestic violence. It was astonishing to observe that many of these reports contained a multiple of new important features that were lacking in the user's understanding of the problem area. An example of such a newly discovered feature was a homosexual relationship. The initial feature set used to train the map, dominantly contained features that were specifically attuned to heterosexual relationships. The result was that police reports describing domestic violence incidents between homosexual men were located in the middle of the non-domestic violence cluster. This is important knowledge for building a highly accurate classifier. Another important feature that was discovered is pepper spray. In about 80% of the domestic violence incidents the perpetrator is a man. By consequence, the feature set contained many terms like "kicking", "stabbing", "maltreating",,.. These violent acts are mostly performed by men. The weapons that are typically used by female offenders were a blank spot in the user's knowledge of the problem area. Again by

investigating these red outliers, we were able to discover that pepper spray is one of the most frequently used weapons of female aggressors. The most important discovered features are displayed in table 3 and 4.

Table 3. Newly discovered features by studying the domestic violence outliers in the ESOM map.

Pepper spray
Homosexual relationship, lesbian relationship
sexual abuse, incest
Alternative spelling of some words (e.g. ex-boyfriend, exboyfriend, ex boyfriend)
Violence terms lacking in the thesaurus: abduction, choke, strangle, etc.
Weapons lacking in the thesaurus: belt, kitchen knife, baseball bat, etc.
Terms referring to persons: partner, fiancée, mistress, concubine, man next door, etc.
Terms referring to relationships: romance, love affair, marriage problems, divorce proceedings, etc.
Reception centers: woman's refuge center, home for battered woman, etc.
Gender of the perpetrator: mostly male
Gender of the victim: mostly female
Age of the perpetrator: mostly older than 18 years and younger than 45 years
Age of the victim: mostly older than 18 years and younger than 45 years
Terms referring to an extra marital affair: I have an another man, lover, I am unfaithful, etc.

Table 4. Newly discovered features by studying the non-domestic violence outliers in the ESOM map.

Places of entertainment: dancing, disco, bar, etc.
Crime locations: on the street, on a bridge, under a viaduct, on a crossing, etc.
Public locations: metro station, bus stop, tram stop, etc.
Reception centers: refugee center, shelter for the homeless, relief center, etc.
Drugs: drug abuse, drug joint, etc.
Addresses of youth institutions, prisons, etc.
Hotel: hotel room, hotel, etc.
Description of suspects origin: Turkish descent, white man, north-African descent, etc.
Description of suspect's body: 1.75 meters tall, 119 centimeters tall, muscular appearance, etc.
Description of suspect's hair: curly haired, blond hair, redhead, etc.
Description of suspect's clothes: black jacket, leather shoes, blue pants, jeans, etc.
Description of suspect's face: beard, moustache, facial hair, etc.
Description of suspect's accent
Unknown person is involved in the crime
Street robbery
Burglary
Car theft
Bicycle theft
Attack by unknown person
Moped theft
Neighborhood quarrel

The reason why some of the non-domestic violence cases, containing one or more of the features of table 4, were located in the middle of the domestic violence cases was because they often contained sentences like "I was walking with *my husband*, when we were suddenly *attacked by an unknown person*". These sentences contain terms that regularly appear in domestic violence reports. By introducing the features presented in table 4, these cases can be better distinguished from the domestic violence cases. As a consequence, more of them will be located in the middle of the non-domestic violence cases in the ESOM map. A new toroidal ESOM map was trained on the dataset based on the refined thesaurus. The resulting map is displayed in Fig. 4.

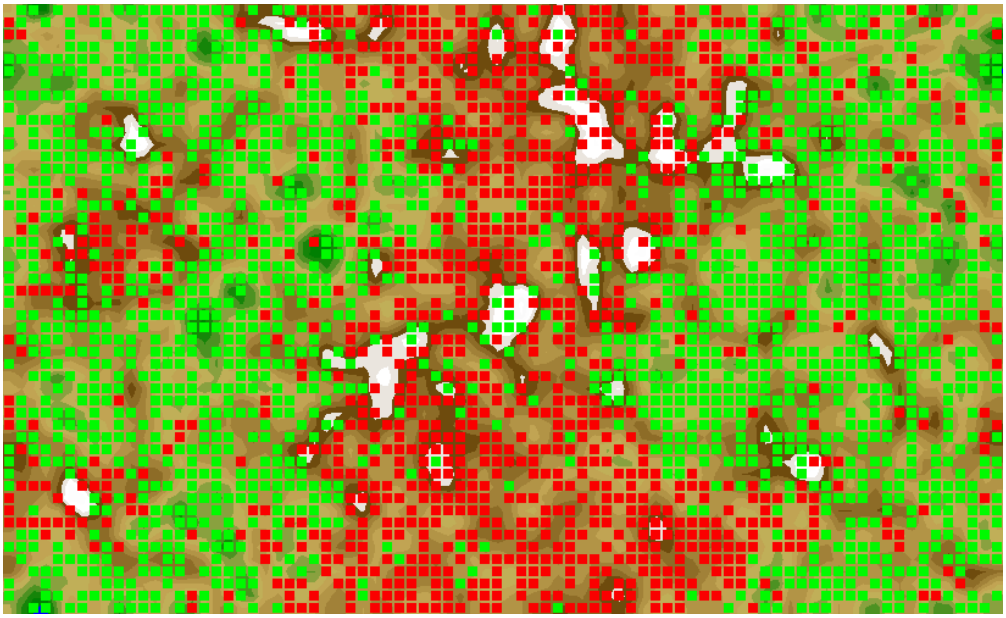

Fig. 4. Toroid ESOM map trained on the categorical dataset with all features

When the ESOM map of Fig. 2 is compared to that of Fig. 3, it is clear that the amount of overlap between the two classes is much lower for the map based on the refined thesaurus. Moreover, the classification accuracy of the SVM, Neural network, Naïve Bayes and kNN classifiers were significantly better after the newly discovered features were added to the thesaurus. For example, for the SVM, the best classification accuracy on the initial dataset was around 83%, while the best classification accuracy on the dataset with the refined thesaurus was around 89%.

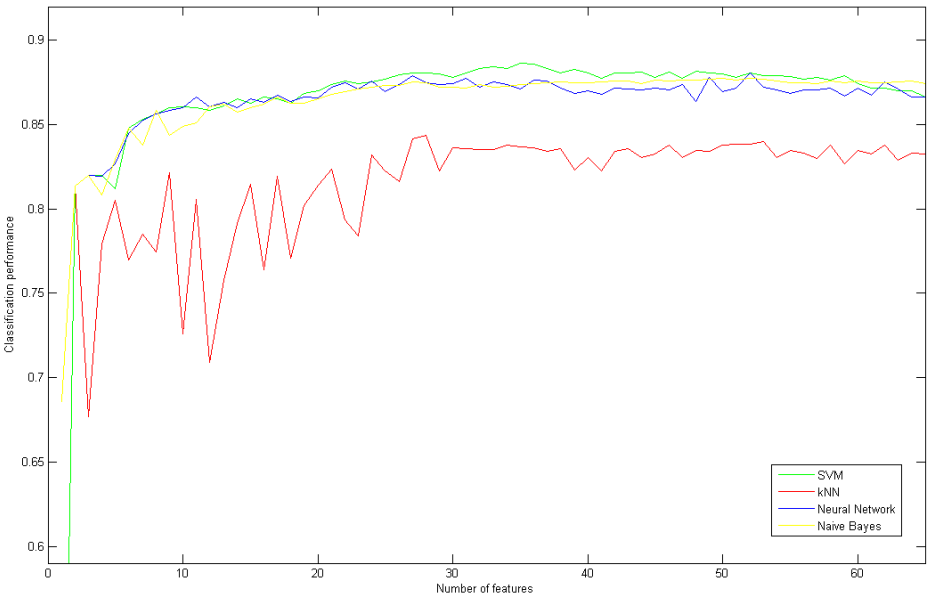

Fig. 5. Classification performance

Another interesting observation was that these outlier neurons often contained cases that cannot be uniquely classified as either domestic or as non-domestic violence on the basis of the definition. For

example, in one case a girl of twelve years old had a relationship of three days with a boy of the same age. After the girl ended the relationship, the boy continued running after her for months. He often went to her home and started banging on the front door, he kept sending messages to her mobile phone, etc. This case was classified as domestic violence by a police officer. However, one may ask whether this is correct. First of all, can we speak of ex-partners if the persons involved had a relationship that lasted only three days? Second, the girl herself was not physically harmed. Only her dad got some minor injuries after he was attacked by the boy. This example makes it clear that it is not always easy to distinguish domestic from non-domestic violence cases. This is often due to the vagueness of the terms contained in the domestic violence definition. Therefore, a pivotal step in the research consisted of clearly defining how broad these terms should be interpreted and clearly demarcating the borderline between domestic and non-domestic violence. After presenting those doubtful cases for which the definition does not provide a unique classification to the board members, responsible for the domestic violence policy, it became clear that the decisive factor for classifying something as domestic violence should be the presence of a dependency relationship between the perpetrator and the victim.

Taking this classification guideline into account, the cases that were outliers in the ESOM map were further investigated. This lead to the discovery of some regularly occurring situations in which there is a clear dependency relationship between the perpetrator and the victim but that were typically classified as non-domestic violence by police officers. An excerpt of these circumstances is listed in table 5.

Table 5. Circumstances in which the offender abuses the dependency relationship with the victim, but that are not recognized by police officers as domestic violence.

Circumstance	Dependency relationship
Lover boys	The victim is in love with the lover boy, who abuses this dependency relationship to make her a prostitute
Extramarital relationship	If the mistress of an adulterer blackmails him, for example by threatening with making their affair known to his wife, the mistress abuses the dependency relationship that exists between her and the man.
Violence between a caretaker and an inhabitant of an institution	If the caretaker threatens or maltreats the inhabitant (for example, a nurse who maltreats an elderly woman in an old folk's home), the latter is often helpless because she depends on the caretaker.
Violence between colleagues	If two colleagues had a relationship and one keeps stalking the other, this is domestic violence between ex-persons.
An ex-boyfriend attacks the new boyfriend	This is considered to be domestic violence because the ex-boyfriend often aims at emotionally hurting his ex-girlfriend.

It was also interesting to observe that some of these outlier cases described incidents that on themselves can not be classified as domestic violence, but might be an early warning indicator for oncoming domestic violence. For example, one incident described an ex-boyfriend who threw a stone through the window of his ex-girlfriend's car. His ex-girlfriend was not threatened, nor physically assaulted by him, because she was not there when the incident took place (the neighbors saw it happen). This isolated incident is not domestic violence; however, it may be a prelude to an escalation of the violence between the two ex-partners. After interviewing a representative number of police officers, it turned out that the majority of them would not assign a domestic violence label to this type of situations. However, according to the board members responsible for the domestic violence policy, this should be classified as domestic violence. This exposed the mismatch between the management's conception of domestic violence and the classification performed by the police officers. The definition employed by the management turned out to be much broader.

To build an optimal classifier, it was necessary to verify whether or not the dataset is stationary, i.e. whether or not there are seasonal influences playing a role in the classification performed by police officers. An ESOM map was trained on the police reports from the year 2007. This map was used to classify the police reports from the four quarters. The results of the nearest neighbor classifiers based on the ESOM map are displayed in table 6 and 7.

Table 6. Classification accuracy of the 1 nearest neighbor classifier applied on the map trained on the dataset of the year 2007.

	Overall accuracy	False Positive Rate	False Negative Rate
Year 2007	88.3%	9.4%	16.0%
1st quarter 2007	92.4%	10.9%	4.2%
2nd quarter 2007	90.6%	8.3%	12.0%
3rd quarter 2007	88.1%	10.1%	15.4%
4th quarter 2007	89.7%	8.9%	13.0%

Table 7. Classification accuracy of the 2 nearest neighbor classifier applied on the map trained on the dataset of the year 2007.

	Overall accuracy	False Positive Rate	False Negative Rate
Year 2007	85.2%	9.9%	23.9%
1st quarter 2007	87.6%	15.5%	9.3%
2nd quarter 2007	86.1%	12.3%	17.3%
3rd quarter 2007	82.5%	13.7%	24.9%
4th quarter 2007	85.9%	10.5%	22.1%

From table 6 and table 7, one may conclude that the overall accuracy of the 1NN classifier based on the ESOM map is better than the overall accuracy of the 2NN classifier based on the same ESOM map. It is clear that there are only minor differences in the classification accuracy on the four maps. Therefore, it is a logical choice to put the datasets of the four quarters of 2007 together in one dataset consisting of 4814 police reports.

An interesting result is the difference in performance of the traditional kNN classifier (around 83%) and the kNN classifier based on the toroidal ESOM map (around 90%). This is due to the topographic map being a model of the data distribution: it forms an approximation of the data manifold, offering interpolating facilities, and it spends more neural hardware at clusters in the data, leading to a modeling of the local density.

It should be noted that more complex classifiers such as the SVM did not perform better than the ESOM, and that the previously developed system were multi-layer perceptrons, which did not provide any insight into the problem (since it is a black-box), and their performance was around 80% only.

Finally, the results were validated on a test set from the year 2006 and the maps and classification performances were similar.

5. Conclusions

Intensive audits of the police databases revealed that many police reports tended to be wrongly classified as domestic or as non-domestic violence. In this paper, it has been shown that the ESOM is an ideal instrument for an in-depth analysis of domestic violence. It was used to discover new features that better distinguish domestic from non-domestic violence cases resulting in higher classification accuracy. Moreover, it proved to be a useful tool for analyzing the domestic violence definition. The mismatch between the management's conception of domestic violence and the classification as performed by police officers was exposed. We found that police officers generally employed a much narrower domestic violence definition in comparison, with the definition employed by the management. Additionally, some regularly occurring situations that were often wrongly classified as non-domestic violence by police officers (e.g. lover boys, etc.) were found using the ESOM. Finally, the ESOM was used to build an accurate, comprehensible and automated classifier.

6. Acknowledgements

The authors would like to thank the police of Amsterdam-Amstelland for providing them with the necessary degrees of freedom to conduct and publish this research. In particular, we are most grateful to Deputy Police Chief Reinder Doeleman and Police Chief Hans Schönfeld for their continued support. Jonas Poelmans is Aspirant of the "Fonds voor Wetenschappelijk Onderzoek – Vlaanderen" (FWO) or Research Foundation – Flanders.

7. References

[1] Keus, R., Kruijff, M.S. (2000) Huiselijk geweld, draaiboek voor de aanpak. Directie Preventie, Jeugd en Sanctiebeleid van de Nederlandse justitie.

[2] Van Dijk, T. (1997) Huiselijk geweld, aard, omvang en hulpverlening (Ministerie van Justitie, Dienst Preventie, Jeugdbescherming en Reclassering, oktober 1997)

[3] Ritter, H. (1999) Non-Euclidean Self-Organizing Maps, pages 97–109. Elsevier, Amsterdam.

[4] Kohonen, T. (1982), "Self-Organized formation of topologically correct feature maps", Biological Cybernetics, Vol. 43, pp 59-69.

[5] Ultsch, A., Moerchen, F. (2005) ESOM-Maps: Tools for clustering, visualization, and classification with Emergent SOM. Technical Report Dept. of Mathematics and Computer Science, University of Marburg, Germany, No. 46

[6] Ultsch, A., Hermann, L. (2005) Architecture of emergent self-organizing maps to reduce projection errors. In Proc. ESANN 2005, PP1-6

[7] Ultsch, A. (2004) Density Estimation and Visualization for Data containing Clusters of unknown Structure. In proc. GfKI 2004 Dortmund, pp 232-239

[8] Ultsch, A. (2003) Maps for visualization of high-dimensional Data Spaces. In proc. WSOM'03, Kyushu, Japan, pp. 225-230

[9] Ultsch, A., Siemon, H.P. (1990) Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis. Proc. Intl. Neural Networks Conf., pp305-308

[10] Van Hulle, M. (2000) Faithful Representations and Topographic Maps from distortion based to information based Self-Organization. Wiley: New York

[11] Peng, H., Long, F., Ding, C. (2005) Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. IEEE Transactions on pattern analysis and machine intelligence, Vol. 27, no. 8.

[12] Argamon, S., Olsen, M. (2006) Toward meaningful Computing, Communications of the ACM, Vol. 49, no. 4

[13] Pednault, E.P.D. (2000) Representation is everything. Communications of the ACM, vol. 43, no. 8

[14] Marchionini, G. (2006) Exploratory search: from finding to understanding. Communications of the ACM, vol. 49, no. 4

[15] Fan, W., Wallace, L., Rich, S., Thang, T. (2006) Tapping the power of text mining. Communications of the ACM, vol. 49, no. 9

[16] http://databionic-esom.sourceforge.net/